Oleic Acid Content in Ground Corn by NIR Spectroscopy with an Indirect Calibration Method

Steve Wright* **and Lisa Hagen**

Pioneer, A DuPont Company, Johnston, Iowa 50131

ABSTRACT: Developing corn hybrids with improved FA profiles is important in providing products with enhanced nutritional characteristics. To support this effort, rapid screening methods are needed to track the various traits of interest. In using NIR methods, calibrations are based on an assumed linear relationship between the concentration of the analyte of interest and the absorbance of the sample. Although this point seems obvious, this linear relationship does not exist when using GC analysis as a reference method for oleic acid content in ground corn kernels. In this case, the GC data provide a relative oleic acid content of the oil and not of the grain from which the NIR spectrum is measured. A method of removing this nonlinearity by modeling the absolute oleic acid content in the grain has been developed. The relative oleic acid content of the oil is then calculated from this predicted absolute oleic acid value, and the total oil content of the grain is predicted from another calibration model. Significant improvement in the model's predictive ability is demonstrated using this two-calibration model.

Paper no. J10578 in *JAOCS 80*, 1163–1167 (December 2003).

KEY WORDS: Calibration development, corn, corn oil, indirect calibration, NIR, nonlinear modeling, oleic acid.

NIR reflectance spectroscopy (NIRS) is a well-established technique for the analysis of quality traits of grain (1–8). NIRS provides a cost-effective high-throughput method to measure these traits and provides the additional advantage of being able to measure multiple characteristics in a single analysis. This ability to measure multiple constituents simultaneously becomes increasingly important, as various quality traits are incorporated into a value-added product.

When developing NIR calibration models, the goal is to identify changes in the spectral absorption that correlate with changes in the analyte's concentration and then to develop a mathematical relationship using the tools of linear algebra. The entire premise of calibration development is based on Beer's law, which defines a simple linear relationship between the absorption spectrum and the chemical composition of a sample. Simply stated, the law claims that when a sample is placed in the optical path of a spectrometer, there is a direct and linear relation between the concentration(s) of its constituent(s) and the amount of energy it absorbs. Mathematically stated: $A_\lambda = \varepsilon_\lambda bC$, where A_λ is the absorbance value of the sample at a specific wavelength λ, ε_λ is the absorptivity coefficient of the constituent at that wavelength, b is the optical path length through the sample, and C is the concentration. Standard calibration methods such as classical least squares (CLS) and inverse least squares (ILS), as well as eigenvector methods such as principal component regression (PCR) and partial least squares (PLS), are all derived from matrix versions of this equation.

Beer's law is based on light absorption in nonscattering media. Agricultural samples such as ground corn seed are highly scattering, and this scattering causes some nonlinear behavior in the spectral absorbancies. To correct for these nonlinear effects, several mathematical preprocessing techniques have been developed (9–15).

Robust NIR calibrations for grain traits require that the models be developed from samples that include the genetic diversity expected in future samples as well as samples from multiple locations and growing years (16). In addition, a linear relationship must exist between the reference chemistry for the analyte of interest and the spectral contribution of that analyte to the total spectrum of the sample being analyzed. Although this point seems obvious, this relationship does not exist when measuring an oil-quality trait such as the oleic acid content in corn seed. To clarify the terminology used, the expression "relative" oleic acid content will refer to the percentage of oleic acid in the oil and the term "absolute" oleic acid content will refer to the actual oleic acid content expressed as a percentage of the total sample (ground corn seed) mass. Because oleic acid is measured in terms of percentage in the oil and not in the grain, and the amount of oil in the grain is not constant, a nonlinear relationship exists between the relative oleic acid content of the oil and the absolute oleic acid content of the grain from which the NIR spectrum is obtained.

For this study, GC analysis was used to provide the relative oleic acid reference data. The oleic acid value from this analysis is expressed as a percentage of the total FA that make up the oil and not the absolute oleic acid content within the sample. To illustrate this nonlinearity, if a corn sample that has a normal oil content of 3.5% were analyzed by GC and determined to have 60% oleic acid, the actual oleic acid content in the grain would be 2.1% [= (60% oleic acid in the oil) × (3.5% oil in the grain)]. If a second sample that was a high-oil pollinator (22% oil) were analyzed and determined to have 35% oleic acid in the oil, the actual oleic acid content in the grain would be 7.7% [= (35% oleic acid in the oil) × (22% oil in the grain)]. From these examples it is clear that fluctuations of oil content in the

---

*To whom correspondence should be addressed at Pioneer Hi-Bred International, 7300 N.W. 62nd St., P.O. Box 1004, Johnston, IA 50131-9412. E-mail: Steven.Wright@Pioneer.com grain will cause a nonlinear relation between the GC reference value and the actual oleic acid content in the grain.

From a spectroscopic point of view, the spectra of the various FA found in corn oil are very similar. The major differences are due to differences in FA chain length and degree of unsaturation. The chain lengths affect the ratio of the CH_3 to CH_2 stretching modes, whereas the degree of unsaturation affects the magnitude of the olefinic C–H stretching and bending modes as well as the magnitude of the C=C stretch. In the NIR region of the spectrum, absorption is due to overtones and combinations of these fundamental absorption frequencies. The problem with developing robust calibrations to measure differences in FA composition in ground corn is that the spectral changes due to changes in FA composition will follow Beer's law with respect to their absolute concentration within the sample and not their composition within the oil (relative oleic acid value).

The genes that affect the oleic acid content also affect the linoleic acid content because both FA are part of the same biochemical pathway. For this reason, there is an inverse relationship between oleic acid and linoleic acid. Since the only difference between the two FA is one double bond, which is not conjugated, one would expect that the major spectral differences would be in the magnitude of the olefinic absorptions. That is, as the oleic acid content of the oil goes up, the magnitude of the olefinic absorption bands goes down. However, all other things being equal, as the oil content in the grain is increased, the absolute magnitude of the olefinic absorption bands is also increased. Conventional modeling can account for this problem to some degree because as the oil content of the grain is increased, the olefinic-to-aliphatic absorption ratio remains unchanged. Nevertheless, conventional modeling techniques such as CLS, ILS, PCR, and PLS are based on linear algebra and work best when dealing with linear relationships and when Beer's law is followed.

EXPERIMENTAL PROCEDURES

Samples and NIR. Corn seed samples used in this study were collected at various Pioneer research stations in the continental United States, Hawaii, and Puerto Rico during the 1997–2000 growing seasons. The samples were diverse in terms of genetics, growing environment, oil content, and oleic acid content. Both transgenic and nontransgenic seeds were included. Nine hundred eighty-four samples were used to construct the calibration models. Prior to NIR scanning, samples were ground through a custom high-throughput self-cleaning grinder. Sample carryover was negligible. NIR spectra of all samples were obtained on an NIR Systems model 6500 scanning monochromator (NIR Systems Inc., Silver Spring, MD) equipped with a sample autochanger. Approximately 3 g of ground corn was placed into a small ring cup (36-mm diameter), and reflectance spectra (log 1/*R*) were recorded from 400 to 2500 nm at 2-nm increments. A total of 16 sample and 8 background scans were averaged for a total analysis time of 1 min. A validation set consisting of 863 samples was used to test both the direct and indirect calibration methods. These samples came from the same growing locations as the calibration set but were from the 2001 growing season.

Reference chemistry. The oleic acid content of the samples was determined using GC. Approximately 5 mL of hexane was added to ~500 mg ground corn seed, and the samples were agitated for 1 min and allowed to sit for 15 min to extract the oil. The extract was then transferred to autosampler vials, and 100 μL of trimethylsulfonium hydroxide in methanol was added to effect methylation. Samples (1 μL) were injected using split injection (30:1) onto a 6890 Plus gas chromatograph (Agilent). FAME were separated on a SUPELCOWAX 10 15 m × 0.25 mm i.d. × 0.25 μm film thickness column and detected using FID. Percent FA composition (16:0–18:3) was determined using Turbochrom 6.1 software (PE Nelson, Wellesley, MA).

The oil content of the samples was determined using ether extraction (17).

Calibration development. NIR calibration models were developed using Infrasoft International's (Springfield, MO) NIRS 3 calibration development software version 3.10. For each data training set (spectra + reference values), 60 calibration models were developed using various scatter corrections, derivatives, and smoothing functions. This process of evaluating a multitude of preprocessing algorithms was made possible by a custom automated program that allows the user to input the data set and have the software calculate the calibration statistics for the 60 different models. Although this automated procedure is useful for evaluating a wide variety of preprocessing algorithms, a calibration developer must thoroughly scrutinize the calibration data when making decisions such as outlier detection and factor selection.

In this study, two different methods were developed to predict the relative oleic acid content in ground corn seed. The direct method used the relative oleic acid reference values (GC analysis) to build the model directly using a PLS algorithm. The spectral data preprocessing used to treat the data was a weighted-multiplicative scatter correction followed by a gap second derivative using a five-point window and finally a five-point smoothing function. Fifteen factors were required to model the relative oleic acid content.

The second indirect calibration method predicts the relative oleic acid content by using two independent calibration equations. The first equation predicts the absolute oleic acid content in the corn, whereas the second equation predicts the total oil content of the grain. The relative oleic acid content is then calculated by dividing the absolute oleic acid prediction by the total oil prediction. The calibration for total oil had been developed prior to this study and consisted of a data set containing 240 samples. The NIR spectra of these samples were obtained as described above. The oil contents of these samples were obtained by ether extraction (17). The spectral preprocessing for this calibration model was standard normal variate and detrending followed by a gap second derivative using a 10-point window. The spectral data were then smoothed with a five-point smoothing function. A total of eight factors were used to model the total oil content. To develop an equation for absolute oleic acid content, there must be a means of measuring the absolute oleic acid content in the corn seed. Since GC analysis provides relative oleic acid content, absolute oleic acid content must be calculated. Absolute oleic acid content was calculated by multiplying the relative oleic acid content from the GC analysis by the predicted oil content from the NIR calibration. A calibration model was then developed using this calculated absolute oleic acid value. The spectral pretreatment used for this model was standard normal variate and detrending followed by a gap second derivative using a five-point window. The spectral data were then smoothed with a five-point smoothing function. A total of nine factors were used to model the absolute oleic acid content. Cross-validation results of these calibration methods were compared. Both methods were then validated against an independent validation set.

RESULTS AND DISCUSSION

Figure 1 illustrates the results of the cross validation of the direct method. The "S" shape observed in the plot is typical of nonlinear systems. It is also clear that there is a considerable amount of scatter at the high-oleic end of the plot. A slope of less than 1 indicates that the variability in the Y-block data (relative oleic acid content) is not fully accounted for in the X-block data (spectra).

Figure 2 presents the cross-validation results for the absolute oleic acid equation. Although the SE of cross validation (SECV), R-squared, and slope have dramatically improved with this equation, no comparison can be made with the direct method because the predicted values from this model are in different units. It is important to note that there appears to be no indication of nonlinear behavior in this model. To make a comparison of these two methods, these absolute values need to be converted to relative oleic acid content. The relative oleic acid value is calculated by dividing the predicted absolute oleic acid content by the predicted oil values from the total oil equation. The data are then plotted against the reference values obtained from the GC analysis. The results of this calculation are shown in Figure 3. It is apparent from this plot that the "S" shape, characteristic of nonlinear systems, has been removed. It is also clear that both R-squared and slope have significantly improved as well as the SECV. In addition, the large amount of scattering at the high-oleic end of the plot has been removed. Although these results are very encouraging, it is more important to test this method on an independent validation set. Figure 4 illustrates the results of the validation samples predicted using the direct method. Once again, we can see the characteristic "S" shape in the plot of the data. The large scatter at the high end of the oleic acid concentration is also apparent. Figure 5 illustrates the results of the validation set predicted by using the indirect method. Clearly, there has been an improvement in the predicted relative oleic acid values using this method. As before, it is clear that the nonlinearity has been removed, as can be seen by the absence of the "S" shape in the plot. The overall scatter at the high end of the oleic acid concentration range also has been reduced. Overall, there has been an improvement in all of the calibration statistics including the slope, R-squared, and the SE of prediction (SEP).

Although the indirect method improves the predictive accuracy of the relative oleic acid content, interesting differences in the SEP-to-SECV ratio were observed between the two methods. As a calibration model becomes very robust, it would be expected that the SEP-to-SECV ratio would approach unity. For the direct method, this ratio was 1.22, and for the indirect method this ratio was 1.36. This difference implies that the indirect method is less robust than the direct method. However, the indirect method relies on two calibrations to calculate its predictive value. Because the oil calibration was developed before

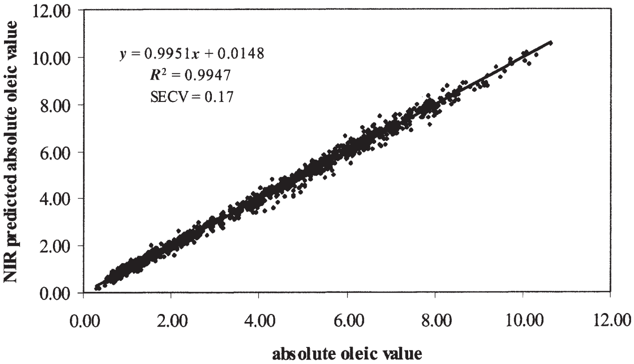

FIG. 2. Cross-validation results of the absolute oleic acid model. For abbreviation see Figure 1. Oleic acid values are % oleic acid by weight in the seed.

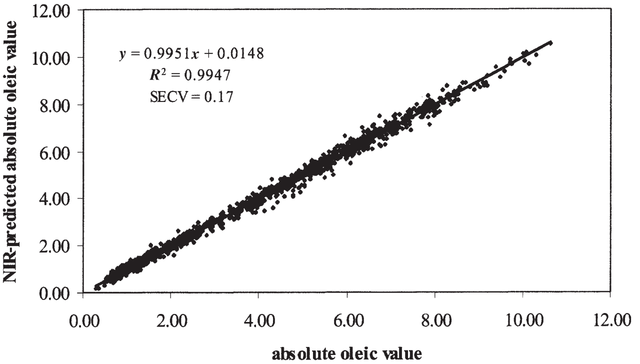

FIG. 1. Cross-validation results of the direct calibration method. SECV, SE of cross validation. Oleic acid values are % oleic acid by weight in the oil.

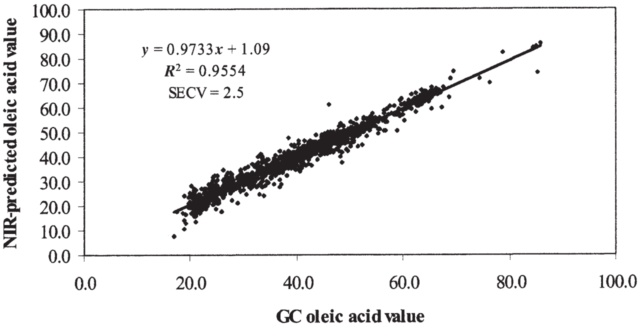

FIG. 3. Cross-validation results of the indirect calibration method. For abbreviation see Figure 1. Oleic acid values are % oleic acid by weight in the oil.

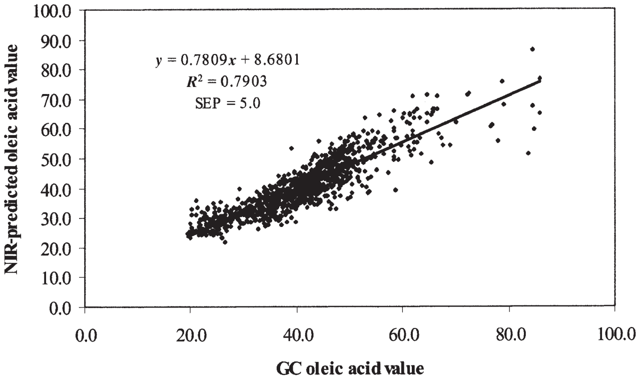

FIG. 4. Validation set results (direct method). SEP, SE of prediction. Oleic acid values are % oleic acid by weight in the oil.

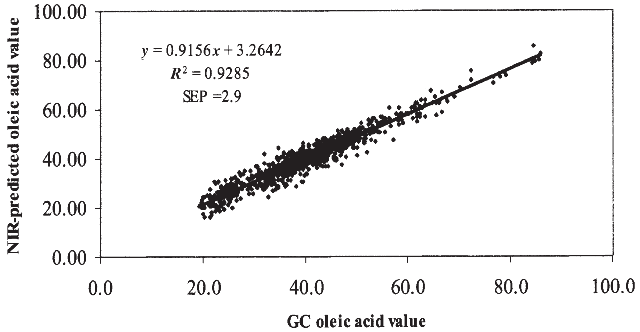

FIG. 6. Validation set results (indirect method with updated oil calibration model). For abbreviation see Figure 4. Oleic acid values are % oleic acid by weight in the oil.

this study began, samples contained in that training set were composed primarily of those that had commodity-type FA profiles [approximately 11% palmitic, 25% oleic, and 60% linoleic acids (18)]. Although the trait of interest (oil) for this calibration model is composed of several FA, the spectral signature for oil is fairly constant because the FA composition in these samples is relatively constant. Thus, the “spectrum” of oil in terms of the olefinic-to-aliphatic absorption ratio remains unchanged among the samples in the calibration training set. Because of this, increased error would be expected when predicting the oil content of samples that have altered FA profiles. To test this theory, the oil calibration was updated with samples that had modified FA compositions. The models for oil and absolute oleic acid were both recalculated. The results of the validation set predicted by using the updated indirect method are shown in Figure 6. Including samples with altered FA improved all of the calibration statistics for the relative oleic acid prediction. The ratio of SEP to SECV also improved to 1.16. It is important to note that since the oil and absolute calibration models are mathematically interrelated, they must be expanded together.

By calculating independent equations that are linear with respect to Beer’s law, the indirect method of calibration removes the nonlinear relation between the GC reference data and the spectral absorbancies. This nonlinear relationship is large in this data set because the oil range covers a full order of magnitude (~2.5–25%). It would be expected that this effect would not be as large for data sets with a more confined range of oil contents. This method of correcting for oil variation has been explored in a previous paper by Daun *et al.* (19).

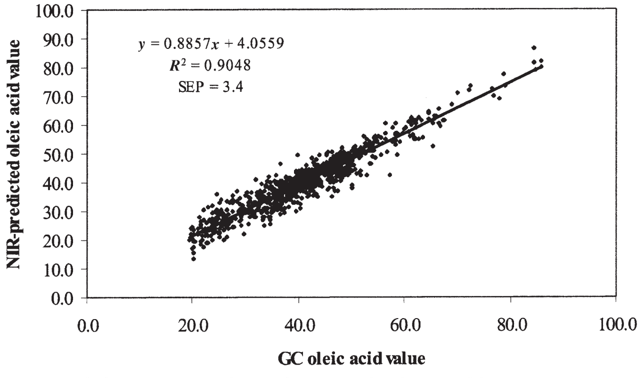

FIG. 5. Validation set results (indirect method). For abbreviation see Figure 4. Oleic acid values are % oleic acid by weight in the oil.

ACKNOWLEDGMENTS

The authors wish to thank Jan Hazebroek and his laboratory staff for providing the reference analysis, Chris Vlahakis for providing the details of the GC analysis, and Roland Welle for writing the automated calibration program.

REFERENCES

1. Velasco, L., B. Pérez-Vich, and J.M. Fernández-Martínez, Nondestructive Screening for Oleic and Linoleic Acid in Single Sunflower Achenes by Near-Infrared Reflectance Spectroscopy, *Crop Sci. 39*:219–222 (1999).
2. Pazdernik, D.L., A.S. Killiam, and J.H. Orf, Analysis of Amino and Fatty Acid Composition in Soybean Seed, Using Near Infrared Reflectance Spectroscopy, *Agron. J. 89*:679–685 (1997).
3. Parrish, F.W., J.P. Madacsi, B.Q. Phillippy, A.G. Wilfred, and S.M. Buco, Determination of Phytic Acid in Cottonseed by Near-Infrared Reflectance Spectroscopy, *J. Agric. Food Chem. 38*:407–409 (1990).
4. Sato, T., I. Uezono, T. Morishita, and T. Tetsuka, Nondestructive Estimation of Fatty Acid Composition in Seeds of *Brassica napus* L. by Near-Infrared Spectroscopy, *J. Am. Oil Chem. Soc. 75*:1877–1881 (1998).
5. Panford, J.A., and J.M. deMan, Determination of Oil Content of Seeds by NIR: Influence of Fatty Acid Composition on Wavelength Selection, *Ibid. 67*:473–482 (1990).
6. Velasco, L., F.D. Goffman, and H.C. Becker, Development of Calibration Equations to Predict Oil Content and Fatty Acid Composition in Brassicaceae Germplasm by Near-Infrared Reflectance Spectroscopy, *Ibid. 76*:25–30 (1999).
7. Velasco, L., B. Pérez-Vich, and J.M. Fernández-Martínez, A Rapid and Simple Approach to Identify Different Sunflower Oil Types by Means of Near-Infrared Reflectance Spectroscopy, *Ibid. 75*:1883–1888 (1998).
8. Velasco, L., J.M. Fernández-Martínez, and A. De Haro, Determination of the Fatty Acid Composition of the Oil in Intact-Seed Mustard by Near-Infrared Reflectance Spectroscopy, *Ibid.* 74:1595–1602 (1997).
9. Faber, N.M., Multivariate Sensitivity for the Interpretation of the Effect of Spectral Pretreatment Methods on Near-Infrared Calibration Model Predictions, *Anal. Chem. 71*:557–565 (1999).
10. Barnes, R.J., M.S. Dhanoa, and S.J. Lister, Standard Normal

Variate Transformation and De-trending of Near-Infrared Diffuse Reflectance Spectra, *Appl. Spectrosc. 43*:772–777 (1989).

11. Geladi, P., D. MacDougall, and H. Martens, Linearization and Scatter-Correction for Near-Infrared Reflectance of Meat, *Ibid. 39*:491–500 (1985).
12. Norris, K.H., and P.C. Williams, Optimization of Mathematical Treatments of Raw Near-Infrared Signal in the Measurement of Protein in Hard Red Spring Wheat. I. Influence of Particle Size, *Cereal Chem. 62*:158–165 (1984).
13. Savitsky, A., and M.J.E. Golay, Smoothing and Differentiation of Data by Simplified Least Squares Procedures, *Anal. Chem. 44*:1627–1639 (1964).
14. Fuller, M.P., G.L. Ritter, and C.S. Draper, Partial Least-Squares Quantitative Analysis of Infrared Spectroscopic Data. Part I: Algorithm Implementation, *Appl. Spectrosc. 42*:217–225 (1988).
15. Chaffin, N.C., and P.R. Griffiths, Role of Scattering Coefficients in Extended Near-Infrared Diffuse Reflection Spectrometry, *Ibid. 52*:218–221 (1998).
16. Shenk, J.S., and M.O. Westerhaus, Population Structuring of Near Infrared Spectra and Modified Partial Least Squares Regression, *Crop Sci. 31*:1548–1555 (1991).
17. AOAC, Fat (crude) or Ether Extraction Animal Feed, *Official Methods of Analysis of AOAC International*, 16th edn., AOAC International, Gaithersburg, MD, 1995, Method 920.39.
18. *Corn: Chemistry and Technology*, edited by S.A. Watson and P.E. Ramstad, American Association of Cereal Chemists, St. Paul, MN, 1994, p. 536.
19. Daun, J.K., K.M. Clear, and P. Williams, Comparison of Three Whole Seed Near-Infrared Analyzers for Measuring Quality Components of Canola Seed, *J. Am. Oil Chem. Soc. 71*: 1063–1068 (1994).

[Received March 4, 2003; accepted September 15, 2003]